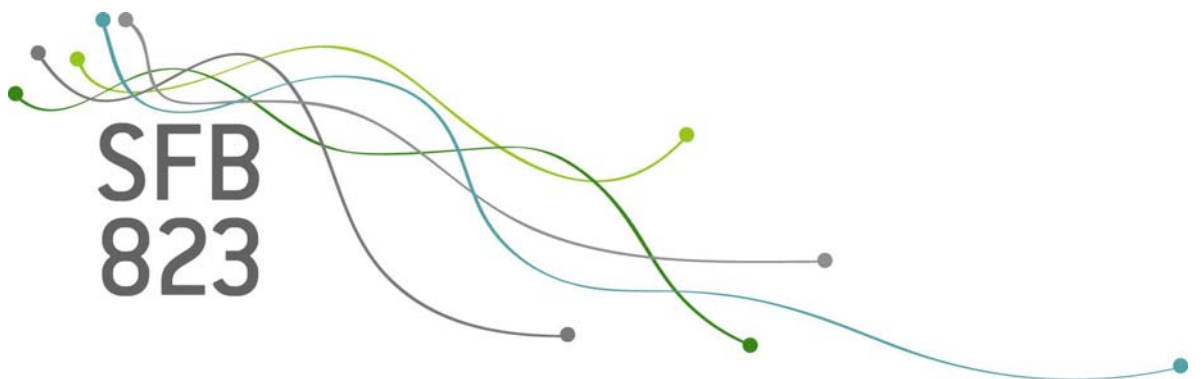# New backtests for unconditional coverage of the expected shortfall

Robert Löser, Dominik Wied, Daniel Ziggel

SFB
823

New Backtests for Unconditional Coverage of the Expected Shortfall[*]

Robert Löser [†]

Technische Universität Dortmund

Dominik Wied [‡]

University of Cologne

Daniel Ziggel [§]

FOM Hochschule für Oekonomie & Management

October 13, 2016

ABSTRACT

We present a new backtest for the unconditional coverage property of the ES. The test statistic is available for finite out-of-sample size which leads to better size and power properties compared to existing tests. Moreover, it can be easily extended to a multivariate test.

Keywords: Model Risk, Multivariate Backtesting, Expected Shortfall.

JEL Classification: C52, C53, C58.

1 Introduction

Since the 1996 Market Risk Amendment to the First Basel Accord, the Value at Risk (VaR) was (and still is) the standard measure for risks of financial investments. Besides, it has become the industry standard not only for banks but also, e.g., for insurance companies (due to Solvency II) or asset managers. However, despite its prevalence, conceptual simplicity and easy interpretation, the VaR has several drawbacks based on fundamental deficiencies. On the one hand, it lacks the desirable property of a coherent risk measure (see Artzner et al., 1999) for non-Gaussian Profit & Loss (P/L) distributions. On the other hand, the VaR does not account for tail risks.

As a consequence, alternative risk measures are of increasing importance and interest with a particular focus on the Expected Shortfall (ES). This is mainly due to the facts that the ES is a coherent risk measure, accounts for tail risks and a consultation paper from the Basel Committee (Basel Committee on Banking Supervision, 2012) opted to replace VaR with ES. However, while calculating ES is not much more difficult in comparison to the VaR, backtesting ES models remains a major challenge (see Yamai and Yoshiba, 2002, 2005; Kerkhof and Melenberg, 2004). While there are several formal VaR-backtests (see, eg. Candelon et al., 2011; Berkowitz et al., 2011; Ziggel et al., 2014, for some recently proposed tests), there are only a few studies dealing with ES-backtests (Berkowitz, 2001; Wong, 2008, 2010; Acerbi and Szekely, 2014). Most recently, Du and Escanciano (2015) proposed some backtests for ES forecasts which are easy to implement. We build upon these recently proposed backtests and present a new backtest for the unconditional coverage (uc) property of the ES. This property is of particular interest for regulators whose focus is mainly on uc backtests.

Our test is based on the so called cumulative violation process. Its main advantage is that the distribution of the test statistic is available for finite out-of-sample size which leads to better size and power properties compared to existing tests. Moreover, it can be easily extended to a multivariate setting. To the best of our knowledge, there is currently

no multivariate backtesting framework for the ES, although this is a highly important task (see Wied et al. (2016) for a discussion concerning multivariate VaR tests).

2 Methodology

In this section, we introduce the notation used throughout the paper, define the desirable properties of VaR and ES models and present our new uni- and multivariate backtests.

2.1 Notation and ES-violation Properties

Let $\{y_t\}_{t=1}^n$ be the observable part of a time series $\{y_t\}_{t\in\mathbb{Z}}$, where y_t represents the return of a bank or an asset at day t. Moreover, let $\{VaR_{t|t-1}(p)\}_{t=1}^n$ be VaR-forecasts at level $p \in (0,1)$, implicitly defined by $P(y_t < -\text{VaR}_{t|t-1}(p)|\mathbb{F}_{t-1}) = p$, where $\mathbb{F}_{t-1}$ denotes the information set up to time $t-1$. The ex-post indicator variable $h_t(p)$ for a given VaR-forecast $VaR_{t|t-1}(p)$ is defined as

$$h_t(p) = \mathbb{1}(y_t < -\text{VaR}_{t|t-1}(p)),$$

where $\mathbb{1}$ denotes the indicator function. In this notation, $h_t(p) = 1$ denotes a VaR-violation.

In this paper we focus on backtesting the ES. Following Du and Escanciano (2015) we define the conditional distribution of y_t given the information set $\mathbb{F}_{t-1}$ as $G_{t|t-1}(\cdot) := G_t(\cdot|\mathbb{F}_{t-1})$. The ES is defined as

$$ES_t := -E(y_t|y_t < -\text{VaR}_{t|t-1}(p), \mathbb{F}_{t-1}) = \frac{1}{p}\int_0^p \text{VaR}_{t|t-1}(u)du.$$

Du and Escanciano (2015) focus on the so called cumulative violations process to test

the correct specification of the ES. The cumulative violation process

$$H_t(p) := \frac{1}{p} \int_0^p h_t(u) du = \frac{1}{p} \int_0^p \mathbb{1}(y_t < -\mathrm{VaR}_{t|t-1}(u)) du$$

$$= \frac{1}{p} \int_0^p \mathbb{1}(G_{t|t-1}(y_t) < u) du = \frac{1}{p} \left(p - G_{t|t-1}(y_t) \right) \mathbb{1}(y_t < -\mathrm{VaR}_{t|t-1}(p))$$

takes values ranging from zero to one if a VaR violation occurs ($h_t(p) = 1$), otherwise it is equal to zero. If $G_{t|t-1}$ is continuous for all t, then $G_{t|t-1}(y_t)$ is uniformly distributed on $[0,1]$. In consequence the expected value at each time t is given by

$$E(H_t(p)) = \frac{p}{2}, \quad \forall t. \tag{1}$$

This is called the uc-property for the ES. As stated in Du and Escanciano (2015), $\{H_t\}$ is unobservable because the true model is unknown and has to be estimated. Therefore some assumptions are necessary:

Assumption 1.

1. *There is a parametric model $G_{t|t-1}(y|\theta_0)$ which specifies the distribution $G_{t|t-1}(y)$ $\forall t$ and $y \leq G_{t|t-1}^{-1}(p)$.*

2. *$G_{t|t-1}(x|\theta)$ is continuously differentiable in θ and $x \in \mathbb{R}$.*

3. *The in-sample of size T is used to estimate the parameter $\theta_0 \in \mathbb{R}^p$ with the consistent estimator $\hat{\theta}_T$.*

These assumptions define our framework and should be fulfilled in most situations.

In the next section we present the uc backtest from Du and Escanciano (2015) and propose our new uc-backtests. The advantage of our tests is that the distribution in finite samples is known.

2.2   New uc-backtest

For testing the hypothesis

$$H_0 : E(H_t(p)) = \frac{p}{2}, \ \forall \ t,$$

$$vs.$$

$$H_1 : \neg H_0 \tag{2}$$

Du and Escanciano (2015) suggest to use a t-test with normal approximation. Since $E(H_t(p)^2) = \frac{p}{3}$ and $\text{Var}(H_t(p)) = \frac{p}{3} - \frac{p^2}{4}$, $\forall t$, the t-test statistic is given by

$$U_{ES} = \sqrt{n} \frac{\overline{H}(p) - p/2}{\sqrt{p(1/3 - p/4)}}$$

with $\overline{H}(p) := \frac{1}{n} \sum_{t=1}^{n} \hat{H}_t(p)$ and $\hat{H}_t(p) = \frac{1}{p}(p - G_{t|t-1}(y_t|\hat{\theta}_T))\mathbb{1}(y_t < G_{t|t-1}^{-1}(p|\hat{\theta}_T))$. Note that $\hat{H}_t(p)$ is used instead of $H_t(p)$ because θ_0 is unknown and has to be estimated by $\hat{\theta}_T$. If $\hat{\theta}_T$ is $\sqrt{T}$-consistent and T increases faster than n (and both tend to infinity) such that $n/T \to 0$, U_{ES} has a standard normal limit distribution.

Now, we consider the case that T tends to infinity but n is fixed and relatively small eg. $n = 250$ or $n = 500$. We simulated this situation $500,000$ times and calculated the statistic U_{ES}. With a density kernel estimation we compare the estimated density function with the standard normal density function. The result is displayed in Figure 1. We see that the simulated distribution is right skewed. Consequently, quantiles for high probabilities obtained from the normal approximation are too small.

Our main contribution is to tackle this problem. For this, we have to reformulate the test hypothesis. We start with the following observation:
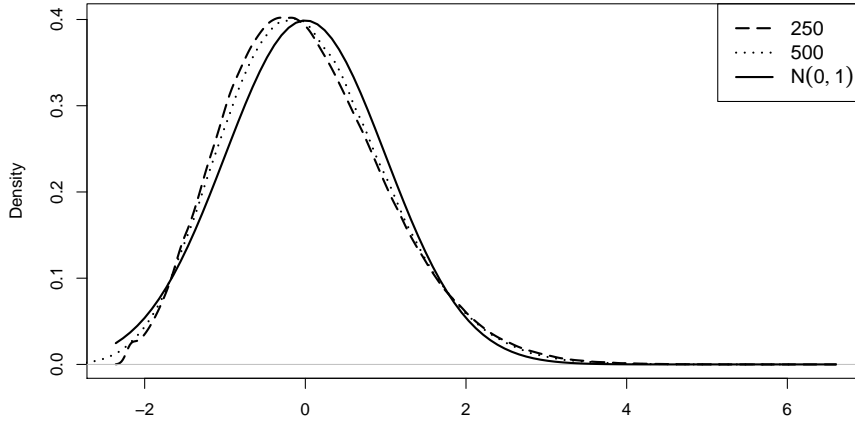
4

Figure 1: Kernel smoothed density function of U_{ES} with out-of-sample-size $n = 250$ and $n = 500$.

Whenever a hit occurs at time t, $G_{t|t-1}(y_t)$ should be uniformly distributed. On the other hand, a hit should occur with probability p if a perfect ES-model is given. Consequently, if the risk model is appropriate, the observable series $\{H_t(p)\}_{t=1}^n$ can be modeled as a series of products of Bernoulli distributed with uniformly distributed random variables. This leads to our test hypotheses:

$$H_0 : H_t(p) = h_t(p) \cdot u_t, \ h_t(p) \sim \mathcal{B}(p) \text{ independent from } u_t \sim \mathcal{U}(0,1), \ \forall t$$

vs.

$$H_1 : \ \neg H_0$$

If H 0 holds, the uc property (1) is obviously fulfilled. Our reformulated hypothesis seems to be stronger than the previous one (2). However, as the calculated expected value and variance in the t-test statistic are based on the reasonable assumption that $E(h_t(u)) = u, \ \forall u \in (0, p)$, it is equivalent to our test hypothesis.

Moreover, no information concerning past values of $H_t(p)$ should be helpful in forecasting hits and their characteristics if the ES-model is correctly specified. Hence, we add the following reasonable assumption:

Assumption 2.

$h_1(p), \ldots, h_n(p), u_1, \ldots, u_n$ *are independent.*

Under this assumption, with the law of total probability, the cumulative distribution function (cdf in the following) of

$$H_n := \sum_{t=1}^{n} H_t(p) = \frac{1}{p} \sum_{t=1}^{n} (p - G_{t|t-1}(y_t|\theta_0)) \mathbb{1}(y_t < G_{t|t-1}^{-1}(p|\theta_0)) \tag{3}$$

is given by

$$F_{H_n}(x) := P(H_n \le x) = \sum_{k=0}^{n} P\left(\sum_{t=1}^{n} h_t(p) = k, \sum_{j=1}^{k} u_j \le x\right).$$

Here, $\sum_{t=1}^{n} h_t(p)$ is binomial distributed with parameter n and p and the distribution of $\sum_{j=1}^{k} u_j$ is the so called Irwin-Hall distribution (Irwin (1927), Hall (1927)) with parameter k and cumulative distribution function

$$\Upsilon_k(x) := \frac{1}{k!} \sum_{j=0}^{\lfloor x \rfloor} (-1)^j \binom{n}{j} (x-j)^{k-1} \mathrm{sgn}(x-j).$$

Thus, the cdf is given by

$$F_{H_n}(x) = \begin{cases} 0 & \text{, if } x < 0 \\ (1-p)^n & \text{, if } x = 0 \\ (1-p)^n + \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \Upsilon_k(x) & \text{, if } x \in (0, n] \\ 1 & \text{, if } x > n. \end{cases}$$

With increasing k, Υ_k is numerically unstable, because $\binom{n}{k}$ takes huge values and/or $p^k(1-p)^{n-k}$ is close to zero. For implementation, it can be useful to use the normal approximation beginning from an upper bound (e.g. $k \ge 20$) $\Upsilon_k(x) \approx \Phi\left(\frac{x-k/2}{\sqrt{k/12}}\right)$.

To demonstrate the usefulness of our test, we simulate $N = 10,000$ series of $\{H_t(p)\}_{t=1}^{n}$

with length $n = 250$ and calculate high quantiles of the simulated $\{H_{n,j}\}_{j=1}^{N}$. The calculated quantiles, the theoretical quantiles from F_{H_n} and the approximation used from Du and Escanciano (2015) are displayed in table 1. The latter is a normal distribution with expected value $np/2$ and variance $n(p/3 - p^2/4)$.

Table 1: Comparison between simulated quantiles and the theoretical quantiles.

quantile $1 - \alpha$	0.95	0.96	0.97	0.98	0.99
simulated $1 - \alpha$- quantile	5.68	5.87	6.11	6.43	6.96
$F_{H_n}^{-1}(1 - \alpha)$	5.67	5.86	6.10	6.43	6.95
$np/2 + \sqrt{(n(p/3 - p^2/4))}\Phi^{-1}(1 - \alpha)$	5.48	5.63	5.81	6.06	6.45

In the following, we assume that there is at least one hit or rather $\hat{H}_n(p) > 0$, with $\hat{H}_n$ defined as in (3) with the estimated parameter $\hat{\theta}_T$ instead of θ_0. We get the continuous conditional cdf

$$F_{H_n|H_n>0}(x) = (1 - (1 - p)^n)^{-1} \sum_{k=1}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \Upsilon_k(x).$$

The conditional cdf is used to define our test statistic

$$S_{UC} := F_{H_n|H_n>0}(\hat{H}_n) = (1 - (1 - p)^n)^{-1} \sum_{k=1}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \Upsilon_k(\hat{H}_n)$$

whose limit distribution is given in the following theorem.

Theorem 3.

Under assumption 1, if n is fixed and $T \to \infty$ then

$$S_{UC} \Big| \left\{ \hat{H}_n > 0 \right\} \xrightarrow{d} \mathcal{U}(0, 1).$$

The proof of this theorem can be found in the appendix.

Thus one would neglect H_0 at level α if $S_{UC} > 1 - \alpha$. We expect that our test has better size properties, because the t-test suffers from the approximation error.

Next, we extend the test to the multivariate framework. Instead of the following approach, one could also simply use the univariate test in combination with a Bonferroni correction. However, this would lead to a less powerful procedure. The price for more power is that we need stronger assumptions, in particular, n must tend to infinity.

With m business lines we define $\mathbf{H}_t := (H_{t,1}(p_1), \ldots, H_{t,m}(p_m))'$ with $H_{t,j}$ the cumulative violation from business line j at day t, $j = 1, \ldots, m$ and $t = 1, \ldots, n$. For simplicity we assume that $p_1 = \ldots = p_m =: p$, but all results can be easily extended for different coverage levels. The test hypothesis is formulated as

$$H_0^m : \mathbf{H}_t = (h_{t,1}(p) \cdot u_{t,1}, \ldots, h_{t,m}(p) \cdot u_{t,m})',$$

$$h_{t,i}(p) \sim \mathcal{B}(p) \text{ independent from } u_{t,i} \sim \mathcal{U}(0,1), \ \forall t, i$$

vs.

$$H_1^m : \ \neg H_0$$

Similarly to assumption 2, we assume independence over time and the same cross-sectional dependence structure at each time point:

Assumption 4.

1. $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_n$ are independent.

2. $Cov(\boldsymbol{H}_1) = \ldots = Cov(\boldsymbol{H}_1) =: \Sigma$.

Clearly, we need that n tends to infinity to estimate Σ consistently. Therefore, it is required that T tends to infinity relatively faster than n. Moreover, we need the following

Assumption 5.

1. $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$.

2. $\exists \, M \in \mathbb{R}| \ \forall t \in \{1, ..., n\} : \ \sup_{\theta \in \Theta} \frac{\partial E(H_t(p)|\mathbb{F}_{t-1})}{\partial \theta} < M$.

8

Assumption 5.1 means that $\hat{\theta}_T$ is $\sqrt{T}$ consistent and the limit distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is bounded in probability. For example, a maximum likelihood estimator with a fixed, rolling or recursive forecasting scheme fulfills this condition as it is shown by Escanciano and Olmo (2010). Assumption 5.2 is more technical and should be fulfilled in nearly any model.

The test statistic is based on the standardized sum of the univariate test statistics. More precisely, the test statistic is given by:

$$S_{UC}^m := \frac{1}{\hat{\sigma}} \sum_{i=1}^m \Phi^{-1}(S_{UC,i}) = \frac{1}{\hat{\sigma}} \sum_{i=1}^m \Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,i}))$$

with $\hat{\sigma}^2$ a consistent estimator for $\mathrm{Var}(\sum_{i=1}^m \Phi^{-1}(S_{UC,i}))$.

Theorem 6.

Under assumption 2, if $T \to \infty$ and $n \to \infty$, $n/T \to 0$, it holds

$$S_{UC}^m | \left\{ \hat{H}_{n,j} > 0, \ \forall j \right\} \xrightarrow{d} \mathcal{N}(0,1),$$

with the variance estimator

$$\hat{\sigma}^2 = \sum_{i=1}^m \sum_{j=1}^m \frac{\sum_{t=1}^n (\hat{H}_{t,i} - \overline{H}_i)(\hat{H}_{t,j} - \overline{H}_j)}{\sqrt{\sum_{t=1}^n (\hat{H}_{t,i} - \overline{H}_i)} \sqrt{\sum_{t=1}^m (\hat{H}_{t,i} - \overline{H}_i)}}.$$

The proof of this theorem can be found in the appendix.

With this theorem, we obtain our multivariate uc-test for backtesting the ES. Conditionally that there is at least one hit in each business line, we reject H_0^m if $S_{UC}^m > u_{1-\alpha}$, with $u_{1-\alpha}$ the $1 - \alpha$ quantile of the normal distribution.

3 Simulation study

In our simulation study we examine the power of our proposed backtest in a controllable but realistic scenario. Also we compare the empirical size and power with the t-Test

proposed by Du and Escanciano (2015). All simulations are computed for a significance level of 5% and we perform 2,000 repetitions for each simulation.

Our test delivers the exact distribution in finite samples if T tends to infinity, but uses basically the same information like the t-Test. Thus we enhanced the size properties, but the size corrected power is the same as the t-Test. So we concentrate on the multivariate setting in this simulation study. In order to extend the univariate AR(1)-GARCH(1,1) model from Du and Escanciano we use a AR(1)-CCC-GARCH(1,1) model with normal- and t-distributed innovations to generate data. The value of the parameters are the same as in Du and Escanciano (2015), thus these are typical für values in empirical application (Du and Escanciano, 2015, p.15). The m-dimensional series $\{Y_t\}_{t=1,...,T+n}$ is generated by:

$$Y_t = 0.05Y_{t-1} + v_t, \ v_t = \sigma_t\epsilon_t, \epsilon_t \sim t(\nu, \Sigma)$$

$$\sigma_t^2 = 0.05\mathbf{1}_m + 0.1\mathbf{I}_m v_{t-1}^2 + 0.85\mathbf{I}_m\sigma_{t-1}^2$$

We use $\nu = \infty$ to generate the in-sample data $\{Y_t\}_{t=1,...,T}$ and fit the model, that leads to normal innovations. If the out of sample is also driven by normal innovations the VaR and ES can be consistently estimated and H_0^m holds. To examine the power of our test in comparison to the Bonferroni corrected t-Test we simulate a structural break at time T, after that point the innovations are t-distributed with $\nu \in \{30, 15, 10, 7\}$ degrees of freedom. The chosen covariance matrix Σ will be fixed with $\Sigma_{ij} = 0.4$ if $i \neq j$ and $\Sigma_{ii} = 1$.

In each of the simulations we estimate $\theta_0 = (\alpha, \omega, \beta, \gamma)' = (0.05, 0.05, 0.1, 0.85)'$ separately for each business line by $\hat{\theta}_{T,j}$ with the well known conditional maximum likelihood estimation with in-sample size $T \in \{500, 2500\}$. Thus we get for each business line j and

each day t the estimated cdf

$$G_{t|t-1,j}(Y_{t,j}|\hat{\theta}_{T,j}) = \Phi(\hat{\epsilon}_{t,j})$$

with

$$\hat{\epsilon}_{t,j} = \frac{Y_{t,j} - \hat{\alpha}_j Y_{t-1,j}}{\hat{\sigma}_{t,j}},$$

and

$$\hat{\sigma}_{t,j} = \hat{\omega}_j + \hat{\beta}_j(\hat{\sigma}_{t-1}\hat{\epsilon}_{t-1,j})^2 + \hat{\gamma}\sigma^2_{t-1,j}.$$

With this cdf we can calculate the estimated cumulative violation for each day t and each business line j:

$$\hat{H}_{t,j}(p) = \frac{1}{p} \left(p - G_{t|t-1,j}(Y_{t,j}|\hat{\theta}_{T,j})) \right) \mathbb{1}(G_{t|t-1,j}(Y_{t,j}|\hat{\theta}_{T,j}) < p).$$

The simulation results are shown in Tables 2 and 3 below.

As shown, our test clearly outperforms the standard t-test which has extremely bad size properties. Moreover, the size adjusted power of our test is significantly better in all cases. The in-sample size of length $T = 500$ seems to be not sufficient for our test. If we increase the in-sample size, the problems concerning the size vanish at a length of $T = 2500$.

Table 2: Simulation results for $m = 10$ and $m = 50$ Business Lines and out of sample size $T = 2500$

	$m = 10$				$m = 50$			
	$n = 250$		$n = 500$		$n = 250$		$n = 500$	
ν	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test
∞	0.05	0.14	0.06	0.12	0.05	0.24	0.06	0.21
30	0.33	0.21	0.53	0.35	0.44	0.28	0.64	0.42
15	0.76	0.51	0.94	0.76	0.85	0.66	0.98	0.87
10	0.93	0.77	1.00	0.96	0.98	0.90	1.00	0.99
7	0.99	0.96	1.00	1.00	1.00	0.98	1.00	1.00

Table 3: Simulation results for $m = 10$ and $m = 50$ Business Lines and out of sample size $T = 500$

	$m = 10$				$m = 50$			
	$n = 250$		$n = 500$		$n = 250$		$n = 500$	
ν	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test	UC_{10}-Test	t-Test
∞	0.12	0.30	0.14	0.36	0.13	0.54	0.15	0.62
30	0.26	0.16	0.43	0.22	0.39	0.19	0.43	0.27
15	0.62	0.38	0.86	0.58	0.80	0.46	0.86	0.65
10	0.89	0.66	0.98	0.86	0.97	0.76	0.98	0.92
7	0.99	0.89	1.00	0.98	1.00	0.96	1.00	1.00

4   Conclusion

We present a new backtest for the unconditional coverage property of the ES. The distribution of the test statistic is available for finite out-of-sample size which leads to better size and power properties compared to existing tests. Moreover, it can be easily extended to a multivariate test. Our test is easy to implement and should be used whenever the in-sample size is large compared to the out-of sample size. To the best of our knowledge this is the first proposed ES backtest for the multivariate setting.

A   Proofs

Proof of Theorem 1

First we prove that for each $t = 1, \ldots, n$ it holds that $\hat{H}_t(p) \overset{p}{\longrightarrow} H_t(p)$:

From Assumption 1 we get $\theta_T \overset{p}{\longrightarrow} \theta_0$ and therefore $G_{t|t-1}(\cdot|\hat{\theta}_T)$ converges uniformly to $G_{t|t-1}(\cdot|\theta_0)$ and of course $G_{t|t-1}^{-1}(\cdot|\hat{\theta}_T)$ converges uniformly to $G_{t|t-1}^{-1}(\cdot|\theta_T)$ since $G_{t|t-1}(\cdot|\hat{\theta}_T)$ is continuous and strictly increasing.

On the one hand we get $\forall x \in \mathbb{R}$

$$\frac{1}{p}(p - G_{t|t-1}(x|\hat{\theta}_T)) \overset{p}{\longrightarrow} \frac{1}{p}(p - G_{t|t-1}(x|\theta_0)).$$

Moreover with $p \in (0,1)$ and y continuously distributed and $0 < \epsilon < 1$

$$P\left(|\mathbb{1}(y < G_{t|t-1}^{-1}(p|\hat{\theta}_T)) - \mathbb{1}(y < G_{t|t-1}^{-1}(p|\theta_0))| > \epsilon\right)$$
$$= P(y \in [\min\{G_{t|t-1}^{-1}(p|\hat{\theta}_T), G_{t|t-1}^{-1}(p|\theta_0)\}, \max\{G_{t|t-1}^{-1}(p|\hat{\theta}_T), G_{t|t-1}^{-1}(p|\theta_0)\}])$$
$$\overset{T \to \infty}{\longrightarrow} P(y = G_{t|t-1}^{-1}(p|\theta_0)) = 0$$

so we get $\mathbb{1}(y < G_{t|t-1}^{-1}(p|\hat{\theta}_T)) \overset{p}{\longrightarrow} \mathbb{1}(y < G_{t|t-1}^{-1}(p|\theta_0))$. Now we use the following Lemma that extends the well known Slutsky-Theorem.

Lemma 7. *If $X_n \overset{p}{\longrightarrow} X$ and $Y_n \overset{p}{\longrightarrow} Y$ then $X_n Y_n \overset{p}{\longrightarrow} XY$.*

Proof

$$P(|X_n Y_n - XY| > \epsilon) = P(|X_n Y_n - X_n Y + X_n Y - XY| > \epsilon)$$
$$\leq P(|X_n(Y_n - Y)| + |Y(X_n - X)| > \epsilon)$$
$$\leq P(|\underbrace{X_n(Y_n - Y)}_{\overset{p}{\longrightarrow} 0}| > \epsilon/2) + P(|\underbrace{Y(X_n - X)}_{\overset{p}{\longrightarrow} 0}| > \epsilon/2) \overset{n \to \infty}{\longrightarrow} 0$$

Using this lemma we get

$$\hat{H}_t(p) = \frac{1}{p}(p - G_{t|t-1}(x|\hat{\theta}_T)) \cdot \mathbb{1}(y < G_{t|t-1}^{-1}(p|\hat{\theta}_T))$$

$$\xrightarrow{p} \frac{1}{p}(p - G_{t|t-1}(x|\theta_0)) \cdot \mathbb{1}(y < G_{t|t-1}^{-1}(p|\theta_0)) = H_t(p)$$

Since n is fixed we immediately obtain

$$\hat{H}_n \xrightarrow{p} H_n,$$

if T tends to infinity. In the second step we show that

$$F_{H_n}\left(\sum_{t=1}^n \hat{H}_t(p) \,\Big|\, H_n > 0\right) \Big| \left\{\sum_{t=1}^n \hat{H}_t(p) > 0\right\} \xrightarrow{d} \mathcal{U}(0,1)$$

Therefore we use another lemma:

Lemma 8.

Let $X_0, X_1, X_2, \ldots$ be continuous random variables on $(\mathbb{R}, \mathbb{F}, P)$ and $X_n \xrightarrow{d} X_0$ and $g : \mathbb{R} \to \mathbb{R}$ absolute continuous and strictly increasing. Then

$$g(X_n)|\{X_n \in B\} \xrightarrow{d} g(X)|\{X \in B\}$$

.

Proof

Since g is absolutely continuous and strictly increasing, $g^{-1} : \mathbb{R} \to \mathbb{R}$ exists. Because X is absolutely continuous it follows by definition of the convergence in distribution and Portmanteau Lemma (see eg. van der Vaart (1998), p.6) for every $B \in \mathbb{F}$

$$P(X_n \in B) \overset{n\to\infty}{\Longrightarrow} P(X \in B).$$

Let $A \subset \mathbb{R}$ and $g^{-1}(A) := \{x \in \mathbb{R} | g(x) \in A\} \in \mathbb{F}$, $B \in \mathbb{F}$ with $P(X_i \in B) > 0 \ \forall i$.

$$
\begin{aligned}
P(g(X_n) \in A | X_n \in B) &= \frac{P(g(X_n) \in A, X_n \in B)}{P(X_n \in B)} \\
&= \frac{P(X_n \in g^{-1}(A), X_n \in B)}{P(X_n \in B)} \\
&= \frac{P(X_n \in \{g^{-1}(A) \cap B\})}{P(X_n \in B)} \\
&\xrightarrow{n \to \infty} \frac{P(X_0 \in \{g^{-1}(A) \cap B\})}{P(X_0 \in B)} \\
&= P(g(X_0) \in A | X_0 \in B)
\end{aligned}
$$

So we get $g(X_n) | \{X_n \in B\} \xrightarrow{d} g(X) | \{X \in B\}$ and Lemma 8 is proved.

With this lemma the proof of Theorem 1 is clear. Per definition it holds

$$
F_{H_n} \left(H_n \middle| H_n > 0 \right) \middle| \left\{ H_n > 0 \right\} \sim \mathcal{U}(0, 1)
$$

and with Lemma 8 it follows easily

$$
F_{H_n} \left(\hat{H}_n \middle| H_n > 0 \right) \middle| \left\{ \hat{H}_n > 0 \right\} \xrightarrow{d} \mathcal{U}(0, 1).
$$

$\square$

Proof of Theorem 2

To prove this theorem, we use a copula theorem from Lindner and Szimayer (2005):

Theorem 9.

Let $(X_n)_{n\in\mathbb{N}}$ and X be m-dimensional random vectors, where $X_n = (X_{n,1}, \ldots, X_{n,m})'$ and $X = (X_1, \ldots, X_m)'$.

Then X_n converges weakly to X as $n \to \infty$, if and only if the margins $X_{n,j}$ converge weakly to X_j as $n \to \infty$ for $j = 1, \ldots, m$, and if the copulas C_n of X_n converge pointwisely to the copula C of X on $RanF_1 \times \ldots \times RanF_m$ as $n \to \infty$, where F_j denotes the distribution function of X_j.

Thus the proof will be done in two steps: First we show that if $T \to \infty$, $n \to \infty$ and $n/T \to 0$ all margins

$$\Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,j}))\Big|\Big\{\hat{H}_{n,j} > 0, \ \forall j\Big\}, \ j = 1, \ldots, m,$$

have standard normal limit distribution. In the second step we show that the copula of $(\Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,1})), \ldots, \Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,m})))'$ converge pointwisely to a Gaussian copula.

If n tends to infinity, we observe $\sup_{x\in\mathbb{R}} |F_{H_n|H_n>0}(x) - F_{H_n}(x)| \longrightarrow 0$ and $P(\hat{H}_{n,j} > 0, \forall j) \longrightarrow 1$. Thus the conditions $\{\hat{H}_{n,j} > 0, \ \forall j\}$ have no effect on the asymptotic behavior and we continue the proof without them. But note that in finite samples this condition is needed to estimate the dependence structure.

$H_{n,j}$ is under Assumption 3 a sum of independent random variables, therefore the central limit theorem holds:

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}} \left|F_{H_n}(x) - \Phi\left(\frac{x - np/2}{\sqrt{n(p/3 - p^2/4)}}\right)\right| = 0.$$

Moreover for $c \in \mathbb{R}$ fixed and $\epsilon \in (0, 0.5)$ it holds:

$$\limsup_{n \to \infty} {}_{x \in \mathbb{R}} \left| F_{H_n}(x + c \cdot n^\epsilon) - \Phi\left(\frac{x + c \cdot n^\epsilon - np/2}{\sqrt{n(p/3 - p^2/4)}}\right) \right| = 0$$

$$\Leftrightarrow \limsup_{n \to \infty} {}_{x \in \mathbb{R}} \left| F_{H_n}(x + c \cdot n^\epsilon) - \Phi\left(\frac{x - np/2}{\sqrt{n(p/3 - p^2/4)}} + \underbrace{\frac{c}{\sqrt{n^{1-2\epsilon}(p/3 - p^2/4)}}}_{\to 0}\right) \right| = 0$$

$$\Leftrightarrow \limsup_{n \to \infty} {}_{x \in \mathbb{R}} \left| F_{H_n}(x + c \cdot n^\epsilon) - \Phi\left(\frac{x - np/2}{\sqrt{n(p/3 - p^2/4)}}\right) \right| = 0.$$

So if $c_n = o_P(\sqrt{n})$ we observe

$$\limsup_{n \to \infty} {}_{x \in \mathbb{R}} \left| F_{H_n}(x) - F_{H_n}(x + c_n) \right| = 0 \ (a.s.)$$

We now consider $d_n := \hat{H}_n - H_n$. Du and Escanciano (2015) mentioned that

$$\frac{1}{\sqrt{n}} d_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E(\hat{H}_t(p) - H_t(p)|\mathbb{F}_{t-1}) + o_P(1),$$

which follows directly from previous results from Escanciano and Olmo (2010). With similar arguments and the mean value theorem we get:

$$
\begin{aligned}
d_n &= \sum_{t=1}^{n}(\hat{H}_t(p) - H_t(p)) \\
&= \sqrt{n}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} E(\hat{H}_t(p) - H_t(p)|\mathbb{F}_{t-1}) + o_P(1)\right] \\
&= \sum_{t=1}^{n}(\hat{\theta}_T - \theta_0) \cdot \left.\frac{\partial E(H_t(p)|\mathbb{F}_{t-1})}{\partial \theta}\right|_{\theta = \tilde{\theta}} + o_P(\sqrt{n}) \\
&= \left[\underbrace{\sqrt{T}(\hat{\theta}_T - \theta_0)}_{O_P(1)} \cdot \frac{n}{\sqrt{T}} \cdot \underbrace{\frac{1}{n}\sum_{t=1}^{n}\left.\frac{\partial E(H_t(p)|\mathbb{F}_{t-1})}{\partial \theta}\right|_{\theta = \tilde{\theta}}}_{B_n} + o_P(\sqrt{n}),\right]
\end{aligned}
$$

with $\tilde{\theta}$ a fixed value between $\hat{\theta}_T$ and θ_0.

If $n/\sqrt{T} = o(\sqrt{n}) \Leftrightarrow n/T \to 0$, we get $d_n = o_P(\sqrt{n})$ and

$$F_{H_n}(\hat{H}_{n_j}) \overset{d}{\longrightarrow} F_{H_n}(H_{n_j}) \sim \mathcal{U}(0,1), \ j=1\ldots,m,$$

and the first step of the proof is completed with the continuous mapping theorem:

$$\Phi^{-1}(F_{H_n}(\hat{H}_{n_j})) \to \Phi^{-1}(F_{H_n}(H_{n_j})) \sim \mathcal{N}(0,1), \ j=1\ldots,m.$$

To complete the proof we determine the dependence structure. Under H_0^m and Assumption 3 we get with the central limit theorem for $\overline{\mathbf{H}} := (\overline{H}_1,\ldots,\overline{H}_m)'$:

$$\widetilde{\mathbf{H}}_n := \sqrt{n}\left(\overline{\mathbf{H}} - \mu\right) \overset{d}{\longrightarrow} \mathcal{N}(0,\Sigma),$$

with $\mu = \frac{p}{2} \cdot 1_m$ and a positive definite matrix $\Sigma \in \mathbb{R}^{m \times m}$.

With Theorem 9 the copula C_n of $\widetilde{\mathbf{H}}_n$ converges pointwisely to a Gaussian copula with a correlation matrix R that corresponds to Σ.

We define $g_n : \mathbb{R}^m \to \mathbb{R}^m$, $g_n((x_1,\ldots,x_m)') \to (g_{1,n}(x_1),\ldots,g_{m,n}(x_n))'$, with

$$g_{i,n}(x_i) = \Phi^{-1}(F_{H_n|H_n>0}(\sqrt{n}x_i + np/2)), \ i=1,\ldots,m.$$

It is easy to see that $g_{i,n}$ is strictly increasing for all $i = 1,\ldots,m$ and $n > 0$ and therefore the distribution of $g_n(\widetilde{\mathbf{H}}_n)$ is also determined by the marginal distributions and the same Copula C_n as before (see eg. Schweizer and Wolff, 1981). Thus we get directly that the copula of $(\Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,1})),\ldots,\Phi^{-1}(F_{H_n|H_n>0}(\hat{H}_{n,m})))'$ is also given by C_n and converges to the Gaussian Copula C. Applying Theorem 9 one more time it holds

$$g_n(\widetilde{\mathbf{H}}_n) \overset{d}{\longrightarrow} \mathcal{N}(0_m, R)$$

18

and

$$1'_m \, g_n(\widetilde{\mathbf{H}}_n) = \sum_{j=1}^{m} \Phi^{-1}(F_{H_n}(\hat{H}_{n,j})) \xrightarrow{d} \mathcal{N}(0, 1'_m R 1_m).$$

As mentioned before, using the conditional cdf and the condition $\{\hat{H}_{n,j} > 0, \; \forall j\}$ does not change the asymptotic behavior. Returning to the conditional case and using the consistent estimator

$$\hat{\sigma}^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\sum_{t=1}^{n}(\hat{H}_{t,i} - \overline{H}_i)(\hat{H}_{t,j} - \overline{H}_j)}{\sqrt{\sum_{t=1}^{n}(\hat{H}_{t,i} - \overline{H}_i)}\sqrt{\sum_{t=1}^{m}(\hat{H}_{t,i} - \overline{H}_i)}},$$

for $1'_m R 1_m$, we derive the limit distribution of the test statistic:

$$\frac{1}{\hat{\sigma}} \sum_{i=1}^{m} \Phi^{-1}(F_{H_n | H_n > 0}(\hat{H}_{n,i})) | \left\{ \hat{H}_{n,j} > 0, \; \forall j \right\} \xrightarrow{d} \mathcal{N}(0, 1),$$

$\square$

References

ACERBI, C. AND B. SZEKELY (2014): "Backtesting expected shortfall," *Risk Magazine*, 32 (7), 1404–1415.

ARTZNER, P., F. DELBAEN, AND D. HEATH (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9(3), 203–228.

BERKOWITZ, J. (2001): "Testing Density Forecasts With Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465–474.

BERKOWITZ, J., P. CHRISTOFFERSEN, AND D. PELLETIER (2011): "Evaluating Value-at-Risk Models with Desk-Level Data," *Management Science*, 57 (12), 2213–2227.

CANDELON, B., G. COLLETAZ, AND S. TOKPAVI (2011): *Backtesting Value-at-Risk: A GMM Duration-Based Test*, vol. 9, Wiley, New York.

DU, Z. AND J. C. ESCANCIANO (2015): "Backtesting Expected Shortfall: Accounting for Tail Risk," *SSRN working paper, online: http://ssrn.com/abstract=2548544.*

ESCANCIANO, J. C. AND J. OLMO (2010): "Backtesting Parametric Value-at-Risk With Estimation Risk," *Journal of Business & Economic Statistics*, 28(1), 36–51.

HALL, P. (1927): "The Distribution of Means for Samples of Size N Drawn from a Population in which the Variate Takes Values Between 0 and 1," *Biometrika*, 9(3/4), 240–245.

IRWIN, J. O. (1927): "On the Frequency Distribution of the Means of Samples from a Population Having any Law of Frequency with Finite Moments, with Special Reference to Pearson's Type II," *Biometrika*, 9(3/4), 225–239.

KERKHOF, J. AND B. MELENBERG (2004): "Journal of Banking & Finance 28 (2004)," *Econometric Theory*, 28, 1845–1865.

LINDNER, A. AND A. SZIMAYER (2005): "A Limit Theorem for Copulas," *Discussion Paper 433, SFB 823.*

SCHWEIZER, B. AND E. F. WOLFF (1981): "On Nonparametric Measures of Dependence for Random Variables," *The Annals of Statistics*, 9 (4), 879–885.

BASEL COMMITTEE ON BANKING SUPERVISION (2012): *Fundamental review of the trading book: A revised market risk framework*, Consultative document, Bank for International Settlements, Basel.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge, New York.

WIED, D., G. WEISS, AND D. ZIGGEL (2016): "Evaluating Value-at-Risk Forecasts: A New Set of Multivariate Backtests," *Journal of Banking & Finance*, 72, 121–132.

WONG, W. K. (2008): "Backtesting trading risk of commercial banks using expected shortfall," *Journal of Banking & Finance*, 32 (7), 1404–1415.

——— (2010): "Backtesting value-at-risk based on tail losses," *Journal of Banking & Finance*, 17, 526–538.

YAMAI, Y. AND T. YOSHIBA (2002): "On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall," *Monetary and Economic Studies*, 20, 57–85.

——— (2005): "Value at Risk versus expected shortfall: a practical perspective," *Journal of Banking and Finance*, 29, 997–1015.

ZIGGEL, D., T. BERENS, G. N. WEISS, AND D. WIED (2014): "A new set of improved Value-at-Risk backtests," *Journal of Banking & Finance*, 48, 29–41.